

B.E. COMPUTER SCIENCE AND ENGINEERING FOURTH YEAR FIRST SEMESTER EXAM 2025

Subject: **DEEP LEARNING**

Time: 3 hours

Full Marks: 100

Answer Question No. 1 (**Compulsory**) and answer **FOUR** questions by choosing ONE from each group

1.

- a. What are the basic differences between supervised and unsupervised learning? - Explain with examples.
- b. Why is *binary cross entropy loss* function used in the *Logistic Regression Model*? Explain.
- c. Compare the *Perceptron training rule* and *Delta rule* used in training single layer neural network - Explain.
- d. Explain why does *Adam* optimization algorithm usually perform better than the *BP algorithm with momentum*?
- e. Is CNN rotation invariant by nature? If yes, explain why? If no, explain how are the rotation variations for images are handled while using CNN for image classification.

4 x 5 = **20 marks**

GROUP A

(Answer Question No. 2 or Question No. 3)

2. a) Considering MSE loss function, derive the weight update rules used for updating weights of a logistic regression model.
- b) For a multivariate linear regression model, if X is an $N \times p$ data matrix with a p -dimensional input vector per row and y is an N -dimensional vector of outputs for N training examples, then show that

$$w^* = (X^T X)^{-1} X^T y$$

where w^* is the best vector of weights.

(10 + 10) = **20 marks**

3. a) What is reinforcement learning? Give an example.
- b) Why is deep learning taking off? Explain.
- c) Consider a dataset in which one data point $(-1, -1)$ belongs to one class, and the other three data points $\{(-1, 1), (1, -1), (1, 1)\}$ belong to the other class. Start with weight parameter values of a linear neuron model at $(0.1, 0.1)$, and use *perceptron rule* for weight updates (incremental mode). Set *learning rate*: $\alpha = 1$ and the bias weight = 0.1. Show calculations up to 2 epochs

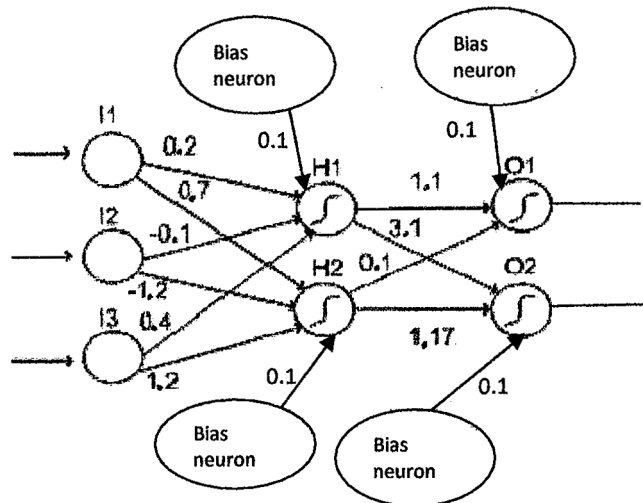
(5+5+10) = **20 marks**

[Turn over

GROUP B

(Answer Question no. 4 or Question no. 5)

4. a) Draw a picture of a biological neuron and show the analogy between a biological neuron and an artificial neuron.
b) Explain why should non-linearity be introduced in multilayer perceptron neural network?
c) Derive the input to hidden layer connection weight update rule for ANN when the hidden nodes use ReLU (rectified linear unit) activation functions and the output nodes use sigmoid activation functions. (5+ 5+10) =20 marks
5. The following is an Artificial Neural Networks, with sigmoid units in the hidden layer and the output layers. The weights have been set arbitrarily between all the units.



Consider that the following training example is submitted to the net (shown in the above figure).

X_1	X_2	X_3	Y_1	Y_2
0.2	0.7	0.3	0.1	0.9

Now show the forward pass to compute responses at H1, H2 and O1, O2. Then use backpropagation (BP) algorithm to find updates for the connection weights. Assume the learning rate $\eta = 0.3$ and momentum = 0.2. You may assume the values of the other hyperparameters if necessary.

5 + 15 = 20 marks

GROUP C

(Answer Question no. 6 or Question no. 7)

6. a) Explain the RMSprop optimization algorithm. How does it differ from Adam optimization algorithm?
b) Explain the advantages of using skip connection in CNN. Mention the name of a standard CNN model where skip connection has been used.
c) In CNN, what are the roles of convolution layer and fully connected layer - explain. 10+ 5+5 = 20 marks
- 7) a) What do you mean by exponentially weighted average? Mention its one application which is not related to deep learning or machine learning.
What is the need for bias correction? In which optimization algorithm it is used?
b) Weight sharing is a key feature of CNNs that differentiates them from other types of neural networks. Explain with a small diagram how weight sharing is done in CNN.
c) What is covariate shift? How can it be reduced?-explain. (5+5) +5+5 = 20 marks

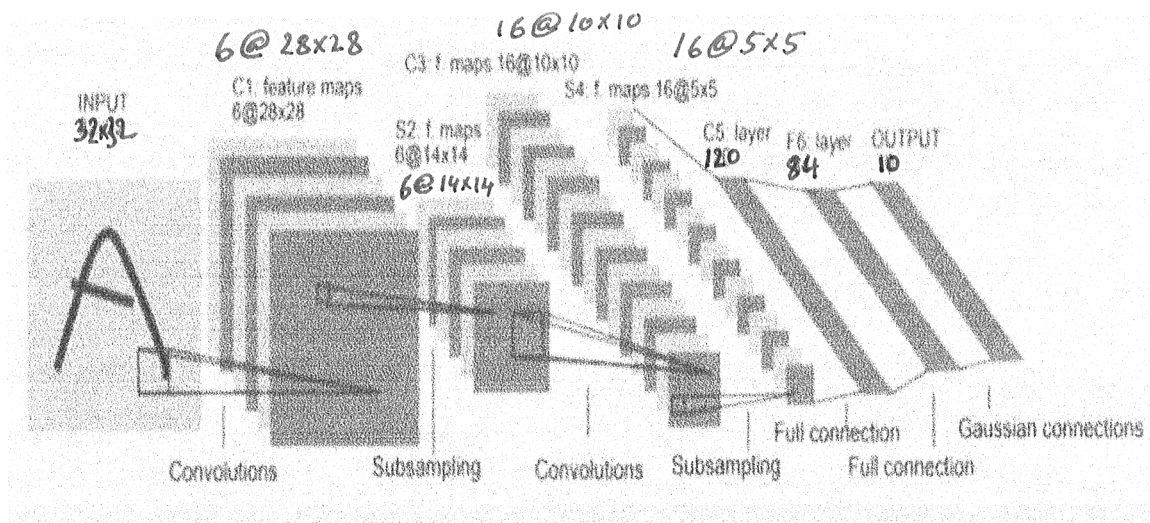
GROUP D

(Answer Question no. 8 or Question no. 9)

8. a) Explain the roles of three gating mechanism used in LSTM network.
 b) What is pre-trained model? What is its use?-Explain.
 c) Explain the architecture of the Unet model. Mention one application where it is commonly used.
 d) What is Word Embedding? Explain the architecture of CBOW (Continuous Bag of Words) word embedding model.

(5 x 4) = 20 marks

9. a) What is training RNN difficult for long input sequences? Explain.
 b) The following is the architecture of Lenet-5. Mention the number of filters, filter size used at each convolutional layer, and pooling size and stride used at each pooling layer of Lenet-5. Calculate the total number of parameters of *Lenet-5*.



- c) In what type of application, word embedding is used and why? Explain briefly the architecture of *Skip Gram Word2Vec* model.

5 + 10 + 5 = 20 marks